1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

United States District Court
Northern District of California

UNITED STATES DISTRICT COURT

NORTHERN DISTRICT OF CALIFORNIA

SAN FRANCISCO DIVISION

IN RE: SOCIAL MEDIA ADOLESCENT
ADDICTION/PERSONAL INJURY
PRODUCTS LIABILITY LITIGATION

Case No. 22-md-03047-YGR   (PHK)

ORDER RE ESI PROTOCOL

Re: Dkts. 352, 534

This MDL has been referred to the undersigned for discovery purposes. *See* Dkt. 426.

Now pending before the Court is the Parties' remaining disputes concerning the terms of a

proposed Electronically Stored Information ("ESI") Protocol. [Dkt. 352; Dkt. 534]. The Court

has carefully reviewed the materials submitted by the Parties and heard oral argument at the

Discovery Management Conference held on January 25, 2024. *See* Dkts. 573, 583.

BACKGROUND

Shortly after this MDL action commenced, the Parties began to meet and confer regarding

the terms of an ESI Protocol. *See* Dkt. 75 at 6; Dkt. 111 at 6-7; Dkt. 164 at 4. On March 3, 2023,

Judge Gonzalez Rogers directed the Parties to submit their remaining disputes as to ESI terms to

Magistrate Judge Hixson by April 14, 2023. [Dkt. 164 at 4]. The Parties thereafter sought

additional time to submit letter briefing on the disputed ESI terms, advising in each instance that

they had been "working diligently to eliminate or narrow disputes over a proposed ESI Protocol,

including by exchanging drafts and through a series of meet-and-confers," but still needed more

time to eliminate or narrow their disputes. [Dkt. 232 at 2; Dkt. 253 at 2; Dkt. 306 at 2; Dkt. 331 at

2]. The Court granted each of these requests, ultimately allowing the Parties until August 10,

1   2023 to submit either a joint proposed ESI Protocol or letter briefing regarding disputes over the

2   ESI terms. *See* Dkts. 254-55, 308, 332.

3   On August 10, 2023, the Parties filed a Joint Statement Re: ESI Protocol setting forth their

4   outstanding disputes over the "purpose and basic framework" for an ESI Order in this case. [Dkt.

5   352]. The Parties attached to the Joint Statement clean and redline versions of their competing

6   proposed ESI orders, as well as prior meet and confer letters exchanged by counsel. Given the

7   manner in which the Parties organized their Joint Statement, it was unclear from the submission

8   both *how many* and *which* ESI terms were still at issue. The Parties attributed the deficiencies to

9   "page constraints." *See* Dkt. 386 at 5.

10   While the ESI dispute was still pending, on September 27, 2023, Magistrate Judge Hixson

11   issued an Order of Recusal in this case. [Dkt. 377]. The case was ultimately reassigned to the

12   undersigned for all further discovery matters. [Dkt. 419].

13   On December 14, 2023, this Court held a Hybrid Discovery Hearing on the pending

14   discovery disputes affecting this case. *See* Dkts. 484, 511. At the hearing, the Court expressed its

15   disappointment as to the excessively large number of disputes regarding ESI issues that remained,

16   as well as the lack of specificity in the Parties' briefing and confusing nature of the competing

17   redlines on those disputes. For that reason, the Court ordered the Parties (including a person most

18   qualified on technical ESI issues from each of their respective ESI vendors) to meet and confer

19   regarding their disputes concerning the entry of an ESI Protocol by December 29, 2023. [Dkt. 503

20   at 4]. The Court further ordered the Parties to promptly submit a joint chart regarding the top ten

21   ESI issues that remained in dispute after the Parties' meet and confer (if any), along with the

22   Parties' respectively proposed language addressing each such ESI issue. *Id.*

23   In accordance with the Court's Order, on January 12, 2024, the Parties filed a Joint Chart

24   of ESI Order Disputes. [Dkt. 534]. In support of the submission, the Parties attested that their

25   counsel and respective ESI consultants "have met and conferred at length in an effort to narrow

26   the parties' disputes." *Id.* at 1. The fourteen-page Joint Chart, which is attached as an exhibit to

27   the filing, identifies the following ESI issues remaining in dispute: (1) Search Methodologies; (2)

28   Validation of Search Methodologies; (3) Collection of Hyperlinks and Production as Families; (4)

United States District Court
Northern District of California

1   Redactions; (5) Databases and Structured Data; (6) Email Threading; (7) Deduplication; (8)

2   System Files; (9) Custodians and Data Sources; and (10) Exception Files. [Dkt. 534-1].

3   　　　　On January 25, 2024, this Court held a Discovery Management Conference ("DMC")

4   regarding the status of discovery issues affecting this MDL. *See* Dkts. 573, 583. At the DMC, the

5   Parties informed the Court that they had reached resolution on Issues 9 and 10. *See* Dkt. 583 at

6   6:7-14. Accordingly, the Court heard extensive oral argument from the Parties as to the eight

7   remaining issues identified by the Parties in their Joint Chart. *See id.* at 6:21-92:8.

8   　　　　　　　　　　　　　　　　**LEGAL STANDARD**

9   　　　　As a general rule, parties in civil litigation "may obtain discovery regarding any

10   nonprivileged matter that is relevant to any party's claim or defense and proportional to the needs

11   of the case." Fed. R. Civ. P. 26(b)(1). The Court has broad authority and discretion to manage

12   this process. *See, e.g.*, *Phillips ex rel. Estates of Byrd v. Gen. Motors Corp.*, 307 F.3d 1206, 1211-

13   12 (9th Cir. 2002) ("[T]he trial court is in the best position to weigh the fairly competing needs

14   and interests of the parties affected by discovery. The unique character of the discovery process

15   requires that the trial court have substantial latitude to fashion protective orders.") (internal

16   quotation marks and alterations omitted); *Hallett v. Morgan*, 296 F.3d 732, 751 (9th Cir. 2002)

17   ("Broad discretion is vested in the trial court to permit or deny discovery[.]") (alteration omitted);

18   *Amarel v. Connell*, 102 F.3d 1494, 1515 (9th Cir. 1996) (recognizing trial courts' "broad

19   discretion to make discovery and evidentiary rulings conducive to the conduct of a fair and orderly

20   trial"); *Asea, Inc. v. S. Pac. Transp. Co.*, 669 F.2d 1242, 1247 (9th Cir. 1981) ("[O]versight of

21   discovery[] is left to the sound discretion of the [court].)"; *see also In re Nat'l Prescription Opiate*

22   *Litig.*, 956 F.3d 838, 841 (6th Cir. 2020) (noting that the Federal Rules of Civil Procedure apply

23   with equal force in multidistrict litigation and observing that "an MDL court has broad discretion

24   to create efficiencies and avoid duplication—of both effort and expenditure—across cases within

25   the MDL").

26   　　　　Here, the Parties have been unable to come to an agreement regarding many of the specific

27   terms of a proposed ESI Protocol. The Court has broad authority and discretion to assist the

28   Parties in resolving ESI disputes. *Phillips*, 307 F.3d at 1211-12; *see Crosby v. Amazon.com, Inc.*,

United States District Court
Northern District of California

3

No. 21-1083-JCC, 2022 WL 522953, at *1 (W.D. Wash. Feb. 22, 2022) ("Because the parties are unable to come to an agreement regarding all terms contained within an ESI discovery agreement, the Court will, in its discretion, assist the parties in doing so."); *Lawson v. Love's Travel Stops & Country Stores, Inc.*, No. 1:17-CV-1266, 2019 WL 5622453, at *7 (M.D. Pa. Oct. 31, 2019) ("As part of our oversight responsibilities in discovery we have the discretion to modify ESI search term parameters.").

Rule 26(f) requires that Parties meet and confer about ESI at the beginning of discovery. Fed. R. Civ. P. 26(f)(3)(C) (The discovery plan must discuss "any issues about disclosure or discovery of [ESI] including the form or forms in which it should be produced."); *see Hardin v. Mendocino Coast Dist. Hosp.*, No. 17-cv-05554-JST (TSH), 2019 WL 4256383, at *1 (N.D. Cal. Sept. 9, 2019) ("Rule 26 makes clear that the scope of ESI production should be addressed at the beginning of discovery."). An ESI protocol must "be specific or it doesn't mean anything." *Id.* at *2.

After review of the Parties' written submissions and arguments of counsel at the DMC, and applying relevant case law governing appropriate ESI standards, the Court addresses each of the eight remaining ESI disputes below.

DISCUSSION

While the Parties presented their disputes as encompassing ten disputes (including Issues 9 and 10 which have since been resolved), each of those disputes actually consist of several disputes wrapped up under the same heading. Where relevant, the Court addresses the disputes and sub-disputes, and as directed at the DMC, further **ORDERS** the Parties (including their respective ESI vendors' technical persons most knowledgeable) to meet and confer to resolve several of the disputes which the Parties appear to still be discussing.

Issue 1: Search Methodologies

This Issue is an example of the Parties combining three disputes into one heading. This is in reality three disputes.

Sub-Issue 1(a): Search Methodologies

The Parties' competing proposed language for the ESI Protocol on this issue are as

1   follows:

Plaintiffs' Proposed ESI Language	Defendants' Proposed ESI Language
8. SEARCH METHODOLOGIES The Parties recognize and agree that each party may use one or more search methodologies to cull, review, and produce responsive, not-privileged potentially responsive or relevant information, including the use of keyword search terms and/or the use of technology assisted review ("TAR"). The parties therefore agree to cooperate in good faith, in an iterative manner, regarding the disclosure and formulation of appropriate search methodologies, as well as transparent and cooperative validation procedures. The parties agree to meet and confer and exchange reasonably necessary information regarding any methodologies that remove relevant or responsive information from human review and withhold such information from production (e.g., search terms) and the document sources to which they will be applied, prior to the application of any such methodologies to any ESI source that may contain potentially relevant or responsive ESI. The Requesting Party may propose modifications, such as additional search terms and culling parameters, or TAR enhancements, for a Producing Party to consider. After meeting and conferring, the parties may enter into additional agreements or protocols relating to the search methodologies they will use to satisfy their discovery obligations. To the extent the Parties are unable to reach agreement on the application of, or procedures for, any search or filtering processes, the Parties shall raise such issues for resolution by the Court or its designee. The Parties recognize that as the litigation evolves, there may be a need to supplement earlier agreed methods or search terms to enhance or improve the identification of potentially relevant ESI.	**SEARCH METHODOLOGIES** The Parties shall adopt reasonable and proportionate methodologies to identify, search, collect, cull, review, and produce ESI as reasonably necessary. The Parties recognize and agree that each party may use one or more methodologies to identify, search, collect, cull, review, and produce responsive and non-privileged ESI. The Parties further recognize that different data sets may implicate different methodologies to identify, search, collect, cull, review, and produce responsive and non-privileged ESI. …

Plaintiffs' proposed language raises the risk of delay and invites continuing, ongoing disputes over methodology and search terms. Further, Plaintiffs' language injects details on the TAR issue (discussed below), and is unnecessary for inclusion in this specific part of the ESI Protocol.

1   In light of the above and keeping in mind the need for efficient and effective discovery (as

2 well as the legal standards for ESI in particular) and in the exercise of the Court's discretion, the

3 Court holds that the language for this portion of the ESI Protocol shall be as follows:

4

5 **SEARCH METHODOLOGIES**

6 The Parties shall adopt reasonable and proportionate methodologies to identify, search, collect, cull, review, and produce ESI as required under applicable legal standards. The Parties recognize and agree that each Party may use one or more methodologies to identify, search, collect, cull, review, and produce responsive and non-privileged ESI, including the use of keyword search terms and/or the use of technology assisted review ("TAR") as discussed further herein. The Parties further recognize that different data sets may implicate different methodologies to identify, search, collect, cull, review, and produce responsive and non-privileged ESI. The Parties therefore agree to meet and confer in good faith regarding any potential disputes over their respective ESI productions.

11

12   **Sub-Issue 1(b): Hit Reports from Search Terms**

13   The Parties' competing proposed language for the ESI Protocol on this issue are as

14 follows:

15

Plaintiffs' Proposed ESI Language	Defendants' Proposed ESI Language
9. HIT REPORTS FOR DISPUTED SEARCH TERMS In the event that a Producing Party claims burden with respect to modified and/or additional search terms proposed by the Requesting Party, the Producing Party will provide a hit report for the terms at issue using industry-standard processing tools, such as NUIX or other similar tools. The Producing Defendant will provide a hit report for each custodian or data source in the deduplicated document collection where the terms were applied, including the following with respect to each proposed or modified search term in the collection: 　　a)　　The number of documents with hits for that term; 　　b)　　The number of unique documents, i.e., documents which do not have hits for any other term; 　　c)　　The number of family members, including the documents with hits, of the documents with hits for that term; and	Key Word Search and Hit Reports. If the Producing Party uses search terms to identify, search, or cull potentially responsive ESI, the Producing Party shall disclose the search terms to the Requesting Party. The Parties shall meet and confer regarding any disputes over the disclosed search terms. In the event that a Producing Party claims burden with respect to modified and/or additional search terms proposed by the Requesting Party, the Producing Party will provide a hit report for the terms at issue using industry-standard processing tools such as NUIX or other similar tools. Hit reports are not predictive of richness, prevalence or responsiveness. A Producing Party need not provide a hit report for any term or set of terms that on its/their face is not reasonably and proportionately tailored to identify information responsive to a particular Request (e.g. entire company name, product and/or service name, a custodian's own name, or similar term). A Producing Party

d) The number of unique family members of the documents with hits for that term. The hit report will also include the total number of documents in the de-duplicated collection against which the search terms were applied, the total number of unique documents containing hits, and the total number of unique family members, including the documents with hits, of the documents with hits. It is understood that the unique hit numbers would be document collection specific, as non-U.S. custodial data will not be comingled with U.S. data for deduplication. The Parties will meet and confer to resolve disagreements over the search terms or their application.	need only provide two rounds of hit reports to a Requesting Party. To the extent the Parties are unable to reach agreement on the application of, or procedures for, any search or filtering processes, the Parties shall raise such issues for resolution by the Court. Nothing in this ESI Order may be construed or interpreted as precluding a Producing Party from performing a review to determine if documents captured by search terms are in fact responsive to the Requesting Party's discovery requests. Similarly, nothing in this ESI Order may be construed or interpreted as precluding a Producing Party from performing, by any means, a privilege review of documents determined to be responsive. Further, nothing in this ESI Order requires or may be relied on by a Party to compel the production of documents captured by any search term that are not responsive to the Requesting Party's request, privileged, or otherwise protected from disclosure.

The Court finds that Plaintiffs' proposed language provides more definite guidance on what is to be expected in hit reports as compared to Defendants' holistic approach. Further, at the DMC, Plaintiffs clarified that their term "unique hit" is unrelated to deduplication and instead refers to "the particular documents that that search term that's at issue will hit as opposed to other search terms that may be are agreed upon. And so it isolates how many unique documents would be lost if you remove that search term from the collection." [Dkt. 583 at 10:3-15]. With that clarification, Defendants' counsel agreed that Defendants' concerns about implicating deduplication here were allayed. *Id.* at 10:17-20.

As to numbers of "family members" and "unique family members," Defendants indicated at the DMC that their ESI tools should generally be able to provide such numbers without undue burden, but the specifics of each Defendant's ESI tool may differ. As the Court directed, if there remain good faith disputes after a Requesting Party reviews a hit report, the Parties shall meet and confer to resolve such disputes, including discussing disclosure of numbers of family members and/or unique family members by the Producing Party in a further hit report if that Producing Party's ESI tool is capable of providing such numbers.

In light of the above and keeping in mind the need for efficient and effective discovery (as

United States District Court
Northern District of California

1 well as the legal standards for ESI in particular) and in the exercise of the Court's discretion, the

2 Court holds that the language for this portion of the ESI Protocol shall be as follows:

3

4 **HIT REPORTS AND SEARCH TERMS**

5 If a Producing Party uses search terms to identify, search, or cull potentially responsive ESI,
   the Producing Party shall disclose the search terms to the Requesting Party. The Parties shall
6 meet and confer regarding any disputes over the disclosed search terms. In the event that a
   Producing Party claims burden with respect to modified and/or additional search terms
7 proposed by a Requesting Party, the Producing Party shall provide a hit report for the terms at
   issue using industry-standard processing tools, such as NUIX or other similar tools. The
8 Producing Party shall provide a hit report for the document collection where the terms were
   applied, including the following with respect to each proposed or modified search term in the
9 collection:

10   a)   The number of documents with hits for that term; and
            b)   The number of unique documents, *i.e.*, documents which do not have hits for
11 any other term.

12   If, after reviewing a hit report from a Producing Party, a Requesting Party so chooses,
   it may reasonably request a further hit report which includes:
13
            c)   The number of family members, including the documents with hits, of the
14 documents with hits for that term; and
            d)   The number of unique family members of the documents with hits for that term.
15
            If the ESI tool for the Producing Party is capable without undue burden of providing
16 the number of family members and unique family members, then the Producing Party shall
   provide such further hit report. The Parties (including the person most knowledgeable about
17 the capabilities of the Producing Party's ESI tool and the Requesting Party's person most
   knowledgeable about technical issues from its ESI service provider) shall meet and confer
18 regarding any disputes over whether the Producing Party's ESI tool has the capability or not to
   provider either number of family members or number of unique family members.
19
   The Parties shall meet and confer to resolve disagreements over the search terms or their
20 application. To the extent the Parties are unable to reach agreement on the application of, or
   procedures for, any search or filtering processes, the Parties shall fully comply with the
21 provisions of this Court's Discovery Standing Order regarding the procedure for raising
   discovery disputes with the Court, including the meet and confer and certification requirements
22 therein.

23

24   **Sub-Issue 1(c): Technology Assisted Review (TAR)**

25   The Parties' competing proposed language for the ESI Protocol on this issue are as

26 follows:

27

28

8

Plaintiffs' Proposed ESI Language	Defendants' Proposed ESI Language
8 fn. 2. TAR The producing party will disclose to the receiving party if they intend to use Technology Assisted Review ("TAR") (including predictive coding or any other form of machine learning) to filter out non-responsive documents. The parties will meet and confer at that time to negotiate a suitable TAR protocol.	**TAR**. A Producing Party may use TAR (technology-assisted review) during the culling and review process, which may be applied in addition to search terms. If a Producing Party uses TAR to filter out or exclude non-responsive documents for production, they shall do the following: ● If a Producing Party chooses to apply both search terms and TAR to a review set, the Producing Party will disclose the search terms. ● Disclose the name of the TAR tool used.

At the DMC, the Parties agreed that their competing language was substantively similar, with the only difference being whether a TAR Protocol would be required or not. Plaintiffs' proposal requires a TAR Protocol to be entered into, but (at least as of the DMC) not all Defendants had chosen a TAR tool or even decided whether or not to use TAR. The Parties agreed that Defendants' proposed language does not preclude the Parties from attempting to negotiate a TAR Protocol (nor does the language require a TAR Protocol, as it is silent and open-ended on the issue). The details of a TAR Protocol are not required for this ESI Protocol and may not be needed as a formal matter. The Parties are free to discuss and negotiate a TAR Protocol, of course, as discovery proceeds.

In light of the above and keeping in mind the need for efficient and effective discovery (as well as the legal standards for ESI in particular) and in the exercise of the Court's discretion, the Court holds that the language for this portion of the ESI Protocol shall be as follows:

> **TAR**. A Producing Party may use TAR (technology-assisted review) during the culling and review process of ESI, which may be applied in addition to search terms. If a Producing Party uses TAR to cull, filter out, or exclude documents from that Party's production, that Producing Party shall do the following:
>
> ● If a Producing Party chooses to apply both search terms and TAR to a review set, the Producing Party will disclose the search terms to all other Party/Parties.
>
> ● Disclose the name of the TAR tool or service used to all other Party/Parties.

United States District Court
Northern District of California

Issue 2: Validation of Search Methodologies

The Parties' competing proposed language for the ESI Protocol on this issue are as follows:

Plaintiffs' Proposed ESI Language	Defendants' Proposed ESI Language
10. VALIDATION OF DEFENDANTS' SEARCH METHODOLOGY AND RESULTS. The Parties shall participate in an iterative and cooperative approach in which the Parties will meet and confer regarding reasonable and appropriate validation procedures and random sampling of Defendants' Documents (both of non-responsive sets and of the entire collection against which search terms were run or TAR or other identification or classification methodology was used), in order to establish that an appropriate level of end-to-end recall (the percentage of responsive Documents in the collection which were identified as responsive by Defendants' methodology) has been achieved and ensure that the Defendants' search methodology was effective and responsive ESI was not being inadvertently omitted. **12. REASSESSMENT** After the completion of the search methodology meet and confer sessions and/or after completion of the selected search process itself, a Producing Party, a Receiving Party, or the Court may perceive the need to reassess a search methodology and/or validation process. In all such circumstances, the Parties will meet and confer to address any issues in a reasonable and timely manner; and the time, cost, and/or other resources expended in connection with ineffective methodologies and/or processes shall be deemed irrelevant to the issues of reasonableness and proportionality for additional efforts required.	**Validation.** The Producing Party shall take reasonable steps to validate its review process (i.e., using quality control measures to determine whether its production is missing relevant ESI or contains substantial amounts of irrelevant ESI) and make any necessary adjustments or corrections to its process. The Parties recognize that different validation processes and standards may be appropriate for distinct document/data sets, and may vary per document/data set and Producing Party. If any Requesting Party, upon a showing of good cause, requests additional information regarding the validation method(s) used by the Producing Party, the Producing Party will disclose the methodology used (statistical sampling methodology, other random sampling methodology, etc.). REASSESSMENT [No provision necessary in ESI Protocol].

In general, the Parties agree that some form of validation of their review procedures is appropriate. The disputes center on how much information to share about validation and the burden required to seek additional information from a Producing Party. The Plaintiffs' proposed

United States District Court
Northern District of California

1 language refers to "end-to-end recall" levels for a production. At the DMC, Defendants' counsel

2 agreed that Defendants will provide such numbers for their respective productions to Plaintiffs.

3 [Dkt. 583 at 21:19-22:4]. The follow-on dispute centered on whether a Producing Party would be

4 required to disclose the parameters used by a Producing Party's ESI tool in the sampling or

5 calculation process to obtain the recall numbers. As indicated at the DMC, the Court directs the

6 Parties to first disclose recall numbers and, if those numbers on their face give rise to reasonable

7 disputes over validation, then the Parties shall meet and confer. During any such meet and

8 confers, the Producing Party is encouraged to disclose the parameters for the tool used to calculate

9 recall levels as part of the good faith attempts to resolve disputes short of motions practice.

10   In light of the above and keeping in mind the need for efficient and effective discovery (as

11 well as the legal standards for ESI in particular) and in the exercise of the Court's discretion, the

12 Court holds that the language for this portion of the ESI Protocol shall be as follows:

13 **Validation.** Each Producing Party shall take reasonable steps to validate its review process
   (*i.e.*, using quality control measures to determine whether its production is missing relevant
14 ESI or contains substantial amounts of irrelevant ESI) and make any necessary adjustments or
   corrections to its process. If, after reviewing a Producing Party's production, a Requesting
15 Party reasonably requests additional information regarding the validation method(s) used by
   the Producing Party, the Producing Party will disclose the level of end-to-end recall (the
16 percentage of responsive Documents in the collection which were identified as responsive by
   that Producing Party's methodology). If there remain disputes between the Parties regarding
17 validation, the Parties shall meet and confer to resolve such disputes in good faith, including a
   reasonable discussion regarding the tool used and the parameters used to obtain or calculate
18 the level of recall.

19

20 **Issue 3: Collection of Hyperlinks and Production as Families**

21   The Parties' competing proposed language for the ESI Protocol on this issue are as

22 follows:

23

Plaintiffs' Proposed ESI Language	Defendants' Proposed ESI Language
APPENDIX 1 **1) Production Components.** … e) Family relationships between emails and their attachments, embedded files or linked internal or non-public documents and their source document, and connected hardcopy documents will be preserved and maintained	1) Production Components … e) Family relationships between emails and their attachments, embedded files and their source document, and connected hardcopy documents will be maintained in production. Attachments should be consecutively

United States District Court
Northern District of California

in production, subject to Paragraph 13 below. Attachments should be consecutively produced with their parent.	produced with their parent.
13) <u>Hyperlinks.</u> Document(s) and/or folder(s) of documents that are hyperlinked inside a responsive document or responsive documents that may be hyperlinks in otherwise non-responsive emails within a Producing Party's custody, possession, or control, including documents on internal or third-party systems or platforms, do not need to be produced in the first instance as part of the same family group as the Document residing at the location to which that hyperlink points, unless a Party can export files hyperlinked documents automatically (e.g., not manual process) during collection using Google Vault,[1] for which Defendants Google, Snap, and TikTok already have licenses for and use in the ordinary course of business for documents residing in the Google environment, and Microsoft Purview eDiscovery Premium solutions,[2] for which Defendant Meta has already has licenses for and uses in the ordinary course of business for documents residing in the Microsoft environment. Following initial production of documents, (i) if there are particular documents containing hyperlinks for unproduced documents in the production, the Receiving Party may submit a list of such documents by Bates number to the Producing Party, and the Producing Party will engage in reasonable efforts to locate the hyperlinked document at that pointed to location and either identify it by Bates number or provide any non-produced and non-privileged documents, and (ii) if there are particular documents in the production that the Receiving Party believes may have been linked to an ESI custodian's emails, the Receiving Party may submit a list of such documents by Bates number to the Producing Party, and the Producing Party will engage in reasonable efforts to locate such custodian's emails and either identify the produced documents by Bates number or provide any non-produced and non-privileged documents. The parties agree to meet and confer to resolve any future or additional issues regarding hyperlinks.	13) <u>Hyperlinks</u>. The Parties agree that given technological limitations, document(s) and/or folder(s) of documents that are hyperlinked inside a responsive document need not be produced in the first instance. If there are particular hyperlinks identified by the Requesting Party in produced documents, the Requesting Party may submit a list of up to 500 hyperlinks to a particular Producing Party for potentially relevant documents by identifying the Bates number and URL or link text for each requested link to the Producing Party, and the Producing Party will engage in reasonable efforts to locate the hyperlinked document at that location and either identify it by Bates number or provide any responsive, non-produced, and non-privileged documents. Nothing in this provision requires production of non-responsive documents.
1 See https://workspaceupdates.googleblog.com/20	[These footnotes are not appropriate for an ESI Order].

23/12/google-vault-export-hyperlinked-drive-content-from-gmail-messages.html ("Starting December 8, 2023, admins can export Drive files hyperlinked in Gmail messages directly in Google Vault."). Three Defendants, Google, Snap, and TikTok, exclusively use Google Suite in the ordinary course of business. 2 See https://learn.microsoft.com/en-us/purview/ediscovery ("You can use eDiscovery tools in Microsoft Purview to search for content in Exchange Online, OneDrive for Business, SharePoint Online, Microsoft Teams, Microsoft 365 Groups, and Viva Engage teams. You can search mailboxes and sites in the same eDiscovery search, and then export the search results."). Defendant Meta uses Office365 which includes Microsoft Purview.	

Plaintiffs' proposal assumes all or most of the Defendants use Google or Microsoft tools, and further assumes that there exist various capabilities of Google and Microsoft tools which, at the DMC, were admitted to be based on certain reading of documentation about those tools and not based on actual knowledge as to their capabilities. At the DMC, counsel for Defendants represented to the Court, for example, that the tools are incapable of backlinking documents with emails. Defendants' proposal allows for manual retrieval of hyperlinked documents upon request from Plaintiffs, which the Court notes is more robust than an automated system (since an automated search system raises questions of reliability and validation, which the Parties dispute as discussed above). Because the number of documents versus documents/emails that have hyperlinks will vary in unknown amounts amongst the Producing Parties, the Court sets no hard numerical limits (recognizing that Defendants proposed 500 initially and thus appear to tacitly admit that 500 is a reasonable limit). But the Court cautions the Parties not to abuse the ability to request hyperlinked documents.

In light of the above and keeping in mind the need for efficient and effective discovery (as well as the legal standards for ESI in particular) and in the exercise of the Court's discretion, the Court holds that the language for this portion of the ESI Protocol shall be as follows:

1   1)   Production Components

2   …

3   e) Family relationships between emails and their attachments, embedded files and their source document, and connected hardcopy documents will be maintained in production. Attachments should be consecutively produced with their parent.

4

5   13) Hyperlinks. Document(s) and/or folder(s) of documents that are hyperlinked inside a responsive document (including hyperlinked inside emails) within a Producing Party's custody, possession, or control, do not need to be produced in the first instance as part of the same family group as the Document residing at the location to which that hyperlink points. If there are particular hyperlinks identified by the Requesting Party in produced documents, the Requesting Party may submit a list of hyperlinks to a particular Producing Party for potentially relevant documents by identifying the Bates number and URL or link text for each requested link to the Producing Party, and the Producing Party will engage in reasonable efforts to locate the hyperlinked document at that location and either identify it by Bates number or provide any responsive, non-produced, and non-privileged documents. The number of hyperlinks a Requesting Party may identify to a Producing Party shall not be excessive and shall be reasonable, proportional to the needs of the case, and not unduly burdensome.

11

12   **Issue 4: Redactions**

13   The Parties' competing proposed language for the ESI Protocol on this issue are as

14   follows:

15

Plaintiffs' Proposed ESI Language	Defendants' Proposed ESI Language
APPENDIX 1 **18) Redactions.** **(**a) The Producing Party may redact (i) information subject to the attorney client privilege or the work product protection (PRIV); (ii) information that cannot be disclosed pursuant to the Stored Communications Act (SCA); (iii) source code subject to separate agreement applicable to production of source code (CODE); (iv) personal identifying information (PII) of the type described in Fed. R. Civ. P. 5.2 that is not otherwise relevant or responsive (e.g., users' ages, years of birth, illnesses, injuries, and medical diagnoses). (b) No redactions for relevance may be made within a produced document or ESI item. If, during the course of discovery, the Parties identify other kinds of information that any Party has a reasonable basis for redacting, the Parties will meet and confer	18) Redactions. (a) The Producing Party may redact personal identifying information (PII), information that cannot be disclosed pursuant to the Stored Communications Act (SCA), source code (CODE), and information subject to the attorney client privilege or the work product protection (PRIV). Redactions shall not be made solely for relevance within a produced document or ESI item. (b) If, during the course of discovery, the Parties identify other kinds of information that any Party has a reasonable basis for redacting, the Parties will meet and confer regarding those kinds of information before such redactions are made. If the Parties cannot agree, they may seek resolution from the Court. (c) The Producing Party will indicate, on the face of the redaction, the asserted reason(s) for the redaction (PII, SCA, CODE, and/or

on a case-by-case basis regarding that information before such redactions are made. If the Parties cannot agree, they may seek resolution from the Court.

(c) The Producing Party will indicate, on the face of the redaction, the asserted reason(s) for the redaction (PII, SCA, CODE, and/or PRIV) and the REDACTION TYPE metadata field shall indicate that the document contains redactions and the reason(s) for the redaction.

(d) Notwithstanding the foregoing, this provision shall not be read to prohibit redactions required under applicable U.S. law or Protective Order.

(e) Where a responsive document contains both redacted and non-redacted content, the Parties shall produce the non-redacted portions of the document and the OCR text corresponding to the non-redacted portions.

(f) Native Redactions. Spreadsheet files requiring redaction, including without limitation Microsoft Excel files, shall be redacted and produced natively. In addition, a Producing Party may natively redact other files that cannot be properly imaged for redaction.

(g) All images of redacted files shall be processed to show and reveal all comments, revision marks, speaker notes, or other user-entered data which are visible in any view of the document in its native application. Where possible, any occurrences of date/time auto-field items, including in headers and footers, will be removed and replaced with the term AUTODATE to prevent the current date from being printed. Email header information (e.g. date, subject line, etc.) should not be redacted unless it is independently privileged. The production of a document in a redacted form does not affect the Parties' obligation to timely assert and substantiate the assertion of privilege over the content in a privilege log. The Parties shall honor reasonable requests for the production of particular redacted documents in other formats where the image is not reasonably usable.

(h) Color. Redacted versions of documents that would have been produced in color in

PRIV) and the REDACTION TYPE metadata field shall indicate that the document contains redactions and the reason(s) for the redaction.

(d) Notwithstanding the foregoing, this provision shall not be read to prohibit redactions permitted under any applicable law or Protective Order.

(e) Where a responsive document contains both redacted and non-redacted content, the Parties shall produce the non-redacted portions of the document and the OCR text corresponding to the non-redacted portions.

(f) Native Redactions. Spreadsheet files requiring redaction, including without limitation Microsoft Excel files, shall be redacted natively. In addition, a Producing Party may natively redact other files that cannot be properly imaged for redaction.

(g) All images of redacted files shall be processed to show all comments, revision marks, speaker notes, or marks made in track changes, unless such material is redacted and marked in accordance with this section.

h) Color. Redacted versions of documents that would have been produced in color in their un-redacted form shall be produced in color as detailed herein.

| their un-redacted form shall be produced in color as detailed herein. | |

At the DMC, the Parties confirmed that they agree on large portions of this section on redactions including the following subsections from Plaintiffs' proposed language: (1) large parts of subparagraph (a) except the definition of personally identifying information (PII) and insertion of relevance/responsiveness; (2) everything in subparagraph (b) (the first sentence of which in Plaintiffs' proposal is the final sentence of Defendants' subparagraph (a)); (3) all of subparagraph (c); (4) all of subparagraph (d) except whether non-US law should be included; (5) all of subparagraph (e); (6) all of subparagraph (f) except the phrase "and produced" before the word "natively;" (7) most of the first sentence of subparagraph (g) except for the phrase "marks made in track changes;" and (8) all of subparagraph (h).

The Parties do not dispute Plaintiffs' proposed subparagraphs (a)(i)-(iii), and instead, primarily dispute the scope of subparagraph (a)(iv) regarding redactions for PII. With regard to PII that can be redacted under subparagraph (a), at the DMC Plaintiffs agreed that certain PII can be redacted, specifically, phone numbers, personal addresses, personal email addresses, the month and day of birth, and driver's license numbers. [Dkt. 583 at 52:2-59:2]. At the DMC, counsel for Meta and the AG Plaintiffs confirmed they will continue to meet and confer on whether other PII, such as information regarding an individual's age, can remain unredacted from Meta's production due to its potential relevance to the AG Plaintiffs' COPPA claims. The Court **ORDERS** Meta (and any other Defendants with views on this issue) and the Plaintiffs to complete their meet and confers on what other PII can be unredacted (such as age, year of birth, or any other data agreed upon) by the deadlines set forth in the Conclusion below.

The Court notes that the Parties are expected to and shall properly designate documents and materials containing unredacted PII at the appropriate level of confidentiality under the Protective Order, particularly with respect to PII of minors, and shall comply with Federal Rule of Civil Procedure 5.2 in all court filings. The Court notes the Rule 5.2 is not an independent basis for redacting documents during or as part of production in discovery and by its terms applies to filings.

1   With regard to subparagraph (d), the Defendants' proposal would allow redactions based

2   on the law of any jurisdiction, including foreign law. At the DMC, counsel for Snap argued that

3   the European Union's General Data Protection Regulation (GDPR) legal regime could be a basis

4   for redaction. Plaintiffs argued that European law should not be binding on or applicable to a

5   proceeding in this Court, which is governed by U.S. law. Snap cited no binding Ninth Circuit or

6   United States Supreme Court law which requires this Court to apply European substantive law

7   with regard to a discovery issue governed under U.S. law. Accordingly, the Court rejects

8   Defendants' language to the extent it would permit redactions to be made based on an assertion

9   that such redactions are required (or permitted) by foreign law, such as the GDPR. No redactions

10   shall be made based on or justified by non-U.S. law.

11   In light of the above and keeping in mind the need for efficient and effective discovery (as

12   well as the legal standards for ESI in particular) and in the exercise of the Court's discretion, the

13   Court holds that the language for this portion of the ESI Protocol shall be as follows:

---

18) Redactions.

(a) A Producing Party may redact (i) information subject to the attorney client privilege or the work product protection (PRIV); (ii) information that cannot be disclosed pursuant to the Stored Communications Act (SCA); (iii) source code subject to separate agreement applicable to production of source code (CODE); (iv) personal identifying information (PII) including phone numbers, personal addresses, personal email addresses, the month and day of birth, driver's license numbers, and other PII agreed to by the Parties (for example, the Parties shall complete their meet and confer on whether or not to redact users' ages and years of birth). In any event, there shall be no redaction of illnesses, injuries, and medical diagnoses. To the extent a document or pleading contains PII, the Parties shall designate such documents at the appropriate Confidentiality level under the Protective Order and shall comply with Fed. R. Civ. P. 5.2 with regard to filings with the Court.

(b) No redactions for relevance may be made within a produced document or ESI item. If, during the course of discovery, the Parties identify other kinds of information that any Party has a reasonable basis for redacting, the Parties will meet and confer on a case-by-case basis regarding that information before such redactions are made. If the Parties cannot agree, they may seek resolution from the Court.

(c) The Producing Party will indicate, on the face of the redaction, the asserted reason(s) for the redaction (PII, SCA, CODE, and/or PRIV) and the REDACTION TYPE metadata field shall indicate that the document contains redactions and the reason(s) for the redaction.

(d) Notwithstanding the foregoing, this provision shall not be read to prohibit redactions permitted under applicable U.S. law or the Protective Order.

(e) Where a responsive document contains both redacted and non-redacted content, the Parties shall produce the non-redacted portions of the document and the OCR text corresponding to

---

1

2

the non-redacted portions.

(f) Native Redactions. Spreadsheet files requiring redaction, including without limitation Microsoft Excel files, shall be redacted and produced natively (unless the Parties agree to production in some other format). In addition, a Producing Party may natively redact other files that cannot be properly imaged for redaction.

3

4

5

6

7

8

9

(g) All images of redacted files shall be processed to show all comments, revision marks, speaker notes, marks made in track changes, or other user-entered data which are visible in a normal view of the document in its native application, unless such material is redacted and marked as redacted in accordance with this section. Where possible, any occurrences of date/time auto-field items, including in headers and footers, will be removed and replaced with the term AUTODATE to prevent the current date from being printed. Email header information (*e.g.*, date and/or subject line) shall not be redacted unless it is independently privileged. The Parties shall honor reasonable requests for the production of particular redacted documents in other formats where the image is not reasonably usable.

(h) Color. Redacted versions of documents that would have been produced in color in their un-redacted form shall be produced in color as detailed herein.

10

11

Issue 5: Databases and Structured Data

12

The Parties' competing proposed language for the ESI Protocol on this issue are as

13

follows:

14

15

Plaintiffs' Proposed ESI Language	Defendants' Proposed ESI Language
APPENDIX 1 **12) Databases and Other Structured Data.** To the extent that responsive information is stored in a database, or database management system, or proprietary system or otherwise maintained by an application, the Producing Party will identify the database and platform to the Requesting Party, and will meet and confer in good faith in an attempt to reach agreement on the data to be produced and the format and scope of the production. The Producing Party will provide information about the databases to facilitate that discussion.	12) Databases and Other Structured Data. To the extent that a Producing Party produces information that is stored in a database upon production, the Producing Party will identify the database or platform to the Requesting Party. The Producing Party shall produce exports and reports in a reasonably usable form. To avoid doubt, information will be considered reasonably usable when produced in CSV format, tab-delimited text format, Microsoft Excel format, or Microsoft Access format. 16 fn.3. Except as separately agreed, the provisions in Sections 6-11 do not apply to non-custodial data like databases, other structured data, or user data.

16

17

18

19

20

21

22

23

24

25

26

27

The Parties generally agree as to this provision with the exception of whether Defendants'

28

proposed footnote 3 should be included. The proposal's use of the term "non-custodial data"

18

1   appears addressed to an issue of unknown scope at this time, namely, whether there is data that is

2   within the custody/possession/control of a Producing Party (as an entity) but is not "custodial" in

3   the sense that there is an identifiable employee or person who has direct responsibility, possession,

4   custody, or control over that data themselves such that there would be an identifiable "custodian"

5   for such data. At the DMC, Defendants admitted that this issue "[d]epend[s] on how their data is

6   stored, and that's going to look very different across defendants." [Dkt. 583 at 71:9-10]. The

7   problem with Defendants' proposal is that it is vaguely and broadly worded to exempt all "non-

8   custodial data" from all of the provisions of Section 6-11 of the ESI Protocol. It is unclear how

9   much data would thereby be exempted from the ESI Protocol, and Defendants do not explain why

10 such exemption is required. Defendants' proposed footnote 3 is rejected.

11   In light of the above and keeping in mind the need for efficient and effective discovery (as

12 well as the legal standards for ESI in particular) and in the exercise of the Court's discretion, the

13 Court holds that the language for this portion of the ESI Protocol shall be as follows:

APPENDIX 1
12) Databases and Other Structured Data.

To the extent that responsive information is stored in a database, or database management system, or proprietary system or application which has is directed to data storage as one of its primary functions, the Producing Party will identify the database and platform to the Requesting Party. The Producing Party shall produce exports and reports about such responsive information stored in such database, where such exports and reports shall be in a reasonably usable form, and information may be produced in CSV format, tab-delimited text format, Microsoft Excel format, or Microsoft Access format. If there are future disputes over the production of information from a database, the Parties shall meet and confer in good faith in an attempt to reach any further agreements (if needed) on the data to be produced and the format and scope of the production. The Producing Party will provide reasonable amounts of information about the databases to facilitate that discussion.

Issue 6: Email Threading

        The Parties' competing proposed language for the ESI Protocol on this issue are as

follows:

Plaintiffs' Proposed ESI Language	Defendants' Proposed ESI Language
15. EMAIL THREADING	11. EMAIL THREADING
Where multiple email messages are part of a	The parties may use analytics technology to

United States District Court
Northern District of California

single chain or "thread," a Party is only required to produce the most inclusive message(s) ("Last In Time Email(s)") and need not produce earlier, lesser inclusive email messages or "thread members," provided that the program being used to perform threading is disclosed and that (i) all participants in any role (Sender, To, CC and BCC) in any email in the thread are participants, although not necessarily in the same roles, in every email in the thread, (ii) all the subject lines of every email in the thread are identical or are identical except for added prefixes "Re:", or "Fwd:", (iii) the entire body of each thread member is included within the Last in Time Email(s), and (iv) any email in the thread with attachment(s) will be separately produced with the "ThreadID" in the metadata. If a thread member contains any additional or altered data that is not contained in the most-inclusive email (including, without limitation, attachments or BCC recipients), it is not a less-inclusive e-mail and must be separately produced. The Producing Party will honor requests to produce earlier chains of emails otherwise excluded by email thread suppression.	identify email threads and need only produce the unique most inclusive copy and related family members and may exclude lesser inclusive copies. Upon reasonable request, the producing party will produce a less inclusive copy.

The Parties do not dispute that, as a general matter, new threads or different threads (such as email threads that stem off of or branch off of a larger/longer email thread) will need to be produced separately. There is not much dispute about the first and last few lines of Plaintiffs' proposed language, which appears to reach the same result as Defendants' proposed language (which is much shorter as shown above).

The Parties dispute here revolves around the bulk of Plaintiffs' proposed language, starting with Plaintiffs' proposed subparagraphs (i)-(iv) which are directed to defining new or different email threads (which would therefore be produced separately from a larger preceding or parallel email thread). Defendants argue that Plaintiffs' attempts to identify what constitute different or new email threads in their proposed subparagraphs (i)-(iv) are proper, given the different ESI and email dethreading (or hyperthreading) tools used by each of the Defendants.

This dispute highlights a deficiency by the Parties in their meet and confers. Throughout all the meet and confers, Defendants argued that Plaintiffs' proposals were "inaccurate" or did not

1   necessarily align with the capabilities of their respective dethreading tools, but Defendants never

2   told Plaintiffs what those dethreading tools are.

3   The purpose of Plaintiffs' proposed language is to put in what Plaintiffs regard as

4   reasonable "safeguards" to ensure that these hyperthreading tools (whatever they are and however

5   they are configured) will not be misused in some unknown way to remove otherwise responsive,

6   substantive, and presumably non-privileged emails from document productions. While Plaintiffs'

7   motives are understandable, Defendants argue that their proposals risk putting Defendants in

8   violation of the ESI Protocol if their tool is simply incapable of performing as phrased in

9   Plaintiffs' language. Unfortunately, Defendants did not propose any alternate language that would

10   address Plaintiffs' concerns or attempt to craft language that would balance both Parties' concerns.

11   Rather, Defendants' proposal unhelpfully omits any alternative proposed guidelines or agreed

12   language on what constitutes a new thread or not.

13   The Court is disappointed by the lack of communication between the Parties on this issue.

14   As indicated at the DMC, the Parties' counsel expressed willingness to meet and confer further

15   and attempt to reach agreement on this issue. The Court therefore **ORDERS** the Parties to

16   complete their meet and confers on this issue by the deadlines set forth in the Conclusion of this

17   Order.

18   In light of the above and keeping in mind the need for efficient and effective discovery (as

19   well as the legal standards for ESI in particular) and in the exercise of the Court's discretion, the

20   Court holds that the language for this portion of the ESI Protocol shall be as follows:

21   **EMAIL THREADING**

22   The Parties may use analytics technology to identify email threads and shall produce the unique
     most inclusive copy and related family members. Where multiple email messages are part of a
23   single chain or "thread," a Party is only required to produce the most inclusive copy of an email
     message(s) ("Last In Time Email(s)") and need not produce earlier, lesser inclusive email
24   messages or "thread members," provided that the tool or software service being used to perform
     threading is disclosed. A Producing Party may exclude from production lesser inclusive copies
25   of the most inclusive email message in the thread, where the entire body of each of those lesser
     inclusive copies is included within the Last in Time Email. The Producing Party will honor
26   reasonable requests to produce lesser inclusive copies of a Last In Time Email or other earlier
     chains of emails otherwise excluded by email thread suppression.

27

28   The Parties shall treat new or different email chains or threads pursuant to further agreement on
     safeguards or guidelines for defining emails threads which reasonably take into account the

	capabilities of the email threading/dethreading/hyperthreading tools used by each Defendant (which shall be disclosed to Plaintiffs).	

Issue 7: Deduplication

The Parties' competing proposed language for the ESI Protocol on this issue are as follows:

Plaintiffs' Proposed ESI Language	Defendants' Proposed ESI Language
14. DEDUPLICATION Each Party shall make reasonable efforts to globally de-duplicate exact duplicate Documents within the Party's ESI data set across all custodial and non-custodial sources at the family level using MD5 hash values or SHA hash values. "Exact duplicate" shall mean bit-for-bit identity of the Document content with exact Hash Value matches as established by commercial software methods. Hash values of emails will be calculated on the concatenated values of at least the following fields: From, To, CC, BCC, Subject, Date Sent, Time Sent, Body, and the hash values of all attachments. The names of all custodians and non-custodial sources who were in possession of a document prior to deduplication will be populated in the ALL CUSTODIANS metadata field. The original file paths of a document prior to deduplication will be populated in the ALL FILE PATHS metadata field.	10. DEDUPLICATION Each Party shall make reasonable efforts to globally de-duplicate exact duplicate Documents within the Party's ESI data set across all custodians at the family level using industry standard deduplication technology– i.e., MD5 hash values or SHA hash values. The parties will not withhold near-duplicates without meeting and conferring on this issue. The names of all custodians, who were identified as custodians for purposes of collection for this matter, who were in possession of a document prior to deduplication will be populated in the ALL CUSTODIANS metadata field.

The Parties do not dispute that, as a general matter, exact copies of the same document need not be produced. The Parties appear to agree that near-duplicates should not be removed from production entirely. Nor does there appear to be dispute over populating the metadata fields to identify custodians of a document prior to deduplication. Rather, there appears to be mistrust and dispute over the tools and technology Defendants intend to use to perform deduplication.

This dispute highlights the same deficiency by the Parties in their meet and confers: lack of communication and failure to share information. Throughout all the meet and confers, Defendants apparently argued that they need flexibility to use their deduplication tools, which use more than just hash values to identify duplicates. But Defendants never told Plaintiffs what those

1   deduplication tools are. Without that minimal identifying information, Plaintiffs are left to guess

2   whether Defendants' tools are somehow error-prone or will inaccurately remove documents from

3   production. As with the preceding issue, Plaintiffs' proposed language was apparently an attempt

4   to put in safeguards, while Defendants proposed no alternative language but instead put in no

5   safeguards.

6   Further, with regard to email, apparently Plaintiffs' proposed language is addressed to the

7   issue that hash value algorithms do not work particularly well with email, and therefore addressed

8   deduplication by comparing common email fields (such as "from," "to," "cc," and the Subject, and

9   the hash value of any attached documents). According to Plaintiffs, Defendants' technical

10   personnel represented that Defendants' deduplication tool is capable of performing this kind of so-

11   called "concatenated analysis." [Dkt. 583 at 87:12-17]. No explanation from Defendants was

12   given as to their failure to reach agreement on this issue, and at least narrow the dispute over

13   deduplication with respect to emails. The fact that Defendants' ESI technical personnel agreed or

14   admitted that their tool is capable of performing in a manner consistent with Plaintiffs' proposal is

15   a strong indicator that Plaintiffs' proposal is not unduly burdensome.

16   The Court is disappointed by the lack of communication between the Parties on this issue.

17   As indicated at the DMC, the Parties' counsel expressed willingness to meet and confer further

18   and attempt to reach agreement on this issue. The Court therefore **ORDERS** the Parties to

19   complete their meet and confers on this issue by the deadlines set forth in the Conclusion of this

20   Order.

21   In light of the above and keeping in mind the need for efficient and effective discovery (as

22   well as the legal standards for ESI in particular) and in the exercise of the Court's discretion, the

23   Court holds that the language for this portion of the ESI Protocol shall be as follows:

24   **DEDUPLICATION**

25   Each Producing Party shall make reasonable efforts to globally deduplicate exact duplicate
     Documents within that Producing Party's ESI data set across all custodial and non-custodial
26   sources at the family level using either MD5 hash values or SHA hash values or any other
     agreed-upon (and disclosed) industry-standard deduplication technology. The Parties shall
27   reach agreement on such other deduplication technology and shall reach agreement on how
     their deduplication tools shall identify exact duplicates of documents in a manner that is
28   consistent with the disclosed tools and technologies a Producing Party is using. The Parties

shall reach agreement on how to identify exact duplicates of emails using industry-standard commercially available software tools or services, which may for example calculate hash values of emails based on concatenated values of agreed-upon email fields and/or hash values of attachments, or which may use any other method the Parties agree upon.

The Parties shall not withhold from production near-duplicates without meeting and conferring on this issue.

The names of all custodians who were either identified as custodians for purposes of collection for this matter (or otherwise known by the Producing Party to have been in possession or custody of a document prior to deduplication) will be populated in the ALL CUSTODIANS metadata field for the produced version of a document that has duplicates removed from production. The original file paths (if any exist) of a document prior to deduplication will be populated in the ALL FILE PATHS metadata field of the produced document.

Issue 8: System Files

The Parties' competing proposed language for the ESI Protocol on this issue are as

follows:

Plaintiffs' Proposed ESI Language	Defendants' Proposed ESI Language
13. SYSTEM FILES	9. SYSTEM FILES
Each Party will use its best efforts to filter out common system files and application executable files using the national software reference library ("NSRL") NIST hash set list. The Parties also may filter out stand-alone files identified as zero bytes in size. Additional culling of file types based on file header information may be applied to the following, provided these files are not known to be otherwise attached, embedded in, or included with an otherwise responsive document, or are not themselves responsive or contain responsive data or are used to interface with users or interact with or access individual or aggregated user data: Application Package File, Backup Files, Batch Files, Binary Disc Image, C++ File Formats, Cascading Style Sheet, Configuration File, Database File, Dictionary Files, Dynamic Link Library, Event Log Files, Executable Files, Hypertext Cascading Stylesheet, Java Archive Files, JavaScript Files, JavaScript Source Code and Class Files, Macintosh Resource Fork Files, Package Manager Files, Program Files, Program Installers, Python Script Files, Shell Script Files, System or Temporary Files, Thumbnail Cache Files, Troff Files, Truetype Font Files, Windows Cabinet File, Windows	Each Party will use its best efforts to filter out common system files and application executable files using the national software reference library ("NSRL") NIST hash set list. The Parties also may filter out stand-alone files identified as zero bytes in size. Additional culling of file types based on file header information may be applied to the following, provided these files are not known to be otherwise attached, embedded in, or included with an otherwise responsive document: Application Package File, Backup Files, Batch Files, Binary Disc Image, C++ File Formats, Cascading Style Sheet, Configuration File, Database File, Dictionary Files, Dynamic Link Library, Event Log Files, Executable Files, Hypertext Cascading Stylesheet, Java Archive Files, JavaScript Files, JavaScript Source Code and Class Files, Macintosh Resource Fork Files, Package Manager Files, Program Files, Program Installers, Python Script Files, Shell Script Files, System or Temporary Files, Thumbnail Cache Files, Troff Files, Truetype Font Files, Windows Cabinet File, Windows Command Files, Windows File Shortcut, Windows Help Files, Windows Metafiles and Enhanced Metafiles, Windows Spool Files, Windows System File.

Command Files, Windows File Shortcut, Windows Help Files, Windows Metafiles and Enhanced Metafiles, Windows Spool Files, Windows System File.	

The Parties only dispute is whether or not to include Plaintiffs' proposed phrase "or are not themselves responsive or contain responsive data or are used to interface with users or interact with or access individual or aggregated user data" which would (if included) require review of files to determine if they meet any of these conditions. That is, the net effect of Plaintiffs' proposed phrase is either to undo the Producing Party's burden of reviewing the listed types of system files (which generally do not contain useful substantive information) or to put the Producing Party in the position of producing system files based on a good faith reason to believe those system files meet the listed conditions (*i.e.* the system files are responsive, contain responsive data, or are used to interface with users, interacts with or access individual or aggregated user data). Defendants noted at the DMC that if this language is included they would likely err on the side of production in order to avoid accusations that they wrongfully withheld system files. Such increased volume of production of system files, if it occurs, would increase the number of files Plaintiffs would have to review after receipt, as a consequence. However, if a Producing Party lacks a good faith reason to believe a system file contains responsive data or otherwise meets the definition in this phrase, the Court expects that Producing Party to use reasonable and good faith efforts to remove such non-substantive system files from production (and avoid needlessly increasing the volume of production).

Plaintiffs requested this language and thus Plaintiffs cannot be heard to complain later about receiving a large volume of system or junk files (and should not be expected to argue for extensions of the case or discovery schedule on this basis). By arguing for this language, Plaintiffs knowingly undertook the risk that they may receive a large volume of system or junk files as a result. Accordingly, the Court will adopt Plaintiffs' proposed phrase.

In light of the above and keeping in mind the need for efficient and effective discovery (as well as the legal standards for ESI in particular) and in the exercise of the Court's discretion, the Court holds that the language for this portion of the ESI Protocol shall be as follows:

SYSTEM FILES

Each Producing Party will use its best efforts to filter out common system files and application executable files using the national software reference library ("NSRL") NIST hash set list. The Parties also may filter out stand-alone files identified as zero bytes in size. Additional culling of file types based on file header information may be applied to the following, provided these files are not known to be otherwise attached to, embedded in, or included with an otherwise responsive document, or are not themselves responsive or contain responsive data or are used to interface with users or interact with or access individual or aggregated user data: Application Package File, Backup Files, Batch Files, Binary Disc Image, C++ File Formats, Cascading Style Sheet, Configuration File, Database File, Dictionary Files, Dynamic Link Library, Event Log Files, Executable Files, Hypertext Cascading Stylesheet, Java Archive Files, JavaScript Files, JavaScript Source Code and Class Files, Macintosh Resource Fork Files, Package Manager Files, Program Files, Program Installers, Python Script Files, Shell Script Files, System or Temporary Files, Thumbnail Cache Files, Troff Files, Truetype Font Files, Windows Cabinet File, Windows Command Files, Windows File Shortcut, Windows Help Files, Windows Metafiles and Enhanced Metafiles, Windows Spool Files, Windows System File.

Issues 9 and 10: Custodians and Data Sources; and Exception Files

The Parties reported to the Court that, between the date of their submission of the Joint Chart of ESI Order Disputes and the date of the DMC, they reached agreement on these issues. As directed in the Conclusion of this Order, the Parties shall include their agreed upon resolutions of any issues in their final Stipulated and [Proposed] ESI Protocol to the Court by the deadline indicated.

CONCLUSION

The Court's rulings on the disputed ESI terms above shall be incorporated by the Parties into a final Stipulated and [Proposed] ESI Protocol to be submitted to the Court. As noted, there remain a handful of issues which the Parties reported they are still meeting and conferring on for which the Court expects the Parties to resolve and include in the ESI Protocol.

Accordingly, the Parties are **ORDERED** to complete all further meet and confers on the still-under-discussion ESI issues on or before **March 15, 2024**. Those meet and confers shall take place by videoconference or in-person, and shall be undertaken in good faith and promptly, in order to finalize the few remaining issues the Parties have still been discussing. As previously directed, in addition to counsel, the meet and confers shall include, from each of the Parties' respective ESI vendors or service providers, a person most knowledgeable about technical issues
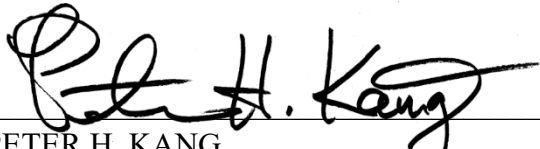
and capabilities for their respective ESI tools.

The Parties are further **ORDERED** to file a final Stipulated and [Proposed] ESI Protocol (and submit a Word version to PHKpo@cand.uscourts.gov) on or before **March 20, 2024** which incorporates all the rulings herein and incorporates the Parties' resolutions of the handful of remaining issues. If the Parties' March 20 submission indicates that the Parties are unable to reach resolution on any of the few remaining open ESI issues, the Court further **ORDERS** lead trial counsel for each Party, all of the attorneys and ESI vendor representatives involved in the prior meet and confers, the CEO of each such ESI vendor, and any other persons involved in managing the discovery issues in this litigation for each of the Parties, to attend mandatory in-person meet and confers on those remaining issues in Courtroom F of the San Francisco courthouse on **March 21, 2024 starting at 8:30 am until 1:00 pm** (and if still not resolved, then continuing on March 22, 2024 starting at 8:30 am until 1:00 pm). Failure to attend and participate in good faith in these meet and confers (if they are needed) shall be grounds for potential sanctions as the Court deems appropriate.

IT IS SO ORDERED.

Dated: February 20, 2024

_____
PETER H. KANG
United States Magistrate Judge